Navy Personnel Research and Development Center

San Diego, CA 92152-6800 TN 88-54 September 1988

An Empirical Bayes Approach to Forecasting Marine Corps Enlisted Personnel Loss Rates

NPRDC TN 88-54

September
1988

An Empirical Bayes Approach to Forecasting Marine Corps Enlisted Personnel Loss Rates

James P. Boyle
Robert M. Holmes, Jr.

Reviewed and released by
Joe Silverman
Director, Manpower Systems Department

Navy Personnel Research and Development Center
San Diego, California 92152-6800

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPRDC TN 88-54	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Navy Personnel Research and Development Center 6b. OFFICE SYMBOL (If applicable) Code 61	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) San Diego, California 92152-6800	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Deputy Chief of Staff for Manpowe 8b. OFFICE SYMBOL (If applicable) MPI-40	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Headquarters, U.S. Marine Corps Washington, DC 20380-0001	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: 63732M PROJECT NO: TASK NO: C0073-03.05 WORK UNIT ACCESSION NO:

11. TITLE (Include Security Classification)

An Empirical Bayes Approach to Forecasting Marine Corps Enlisted Personnel Loss Rates

12. PERSONAL AUTHOR(S)
Boyle, James P. and Holmes, Robert M., Jr.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM Nov 87 TO 5 88	1988 July	30

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Forecasting, loss rates, empirical Bayes, least squares
14	01		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes an empirical Bayes approach to forecasting Marine Corps enlisted personnel loss rates. Particular emphasis is placed on comparing the accuracy of empirical Bayes forecasts with the accuracy of least squares forecasts.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL James P. Boyle	22b. TELEPHONE (Include Area Code) (619) 553-8025 22c. OFFICE SYMBOL Code 61

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

FOREWORD

This research was conducted under program element 63732M (Advanced Manpower/Training Systems), work unit number C0073-03.05, sponsored by the Deputy Chief of Staff for Manpower (MPI-40). The objective of this task was to explore alternative methods for forecasting Marine Corps enlisted personnel loss rates.

This report describes an empirical Bayes approach to forecasting Marine Corps enlisted personnel loss rates. The context is a simple regression model of quarterly loss rates with time as the independent variable. Particular emphasis is placed on comparing the accuracy of forecasts based on empirical Bayes estimates of parameters with the accuracy of forecasts based on standard least squares estimates.

JOE SILVERMAN
Director
Manpower Systems Department

SUMMARY

This report describes an empirical Bayes approach to forecasting Marine Corps enlisted personnel loss rates. The context is a simple time series regression model. The report emphasizes a comparison of the accuracy of forecasts based on empirical Bayes estimates of parameters and the accuracy of forecasts based on least squares estimates of parameters. Additionally, an enlarged class of estimators is developed, the so-called double F empirical Bayes estimators. The performance of the estimators is compared to the performance of the standard single F empirical Bayes estimators. All estimators are derived in Appendix A.

The test data consists of 24 quarterly observations (FY81 through FY86) on end-of-active service (EAS) loss rates for the entire enlisted Marine Corps and three representative occupational fields. For each series and each set of parameter estimates, a mean square error of forecasts is computed to assess forecasting accuracy. Smaller mean square errors correspond to more accurate forecasts and larger mean square errors correspond to less accurate forecasts. All relevant accuracy measures are tabulated.

Two conclusions can be drawn from this investigation. First, the empirical Bayes estimators outperformed the least squares estimators. Second, the double F estimators were not consistently more accurate than the single F estimators.

CONTENTS

INTRODUCTION

Least squares regression is the most common method for estimating parameters in a linear model. Recently, the popularity of least squares has been challenged by a new set of techniques known as "Empirical Bayes" methods. These methods furnish estimates that have been shown to be superior to the standard least squares estimates, at least in terms of smaller mean square errors.

To date, the principal papers on the empirical Bayes estimator (Efron & Morris, 1972, 1973, 1975, 1977; Carter & Rolph, 1974; Fay & Herriot, 1979) have presented the method as a weighted average of the unrestricted least squares estimator and a restricted least squares estimator. Equivalently, it shrinks the unrestricted estimates towards restricted estimates. Judge and Bock (1978), and later Casella (1985), noted that the amount of shrinkage is proportional to an F-statistic testing the hypothesis that the parameters satisfy the restrictions.

This report describes a comparison of forecasts generated by least squares estimators and empirical Bayes estimators of parameters in the context of a simple linear model of Marine Corps enlisted personnel loss rates. The comparison tests the validity of the theoretical findings in the literature. Special attention is paid to comparing the performance of single F empirical Bayes estimators to double F empirical Bayes estimators. These estimators are derived in Appendix A.

APPROACH

Model Specification

The test data used for this study consists of 24 quarterly observations (FY81 through FY86) on end-of-active service (EAS) loss rates for the entire enlisted Marine Corps (ALLMAR) and 3 occupational fields (OccF): OccF3 (Infantry), OccF25 (Operational Communications), and OccF30 (Supply Administration and Operations)[1]. An "EAS loss rate" is defined as the fraction of a personnel inventory at the beginning of a quarter that is "at risk" (those with service contracts expiring sometime during the quarter), and that leave active duty during that quarter. Put differently, the loss rate is the fraction of the "at risk" inventory not extending their contracts or reenlisting. See Table B-1 of Appendix B for the actual historical loss rate values.

Figures 1a-1d display plots of the four loss rate series. An inspection of these plots reveals trends across years, as well as seasonal variation within years. This behavior is captured by the following linear model:

$$Y_{tj} = \alpha_j + \beta_j (t - \bar{t}) + \varepsilon_{tj} \quad (1)$$

where Y_{tj} is the observed loss rate for year t and quarter j. The error terms, ε_{tj}'s, are assumed to be uncorrelated with constant variance. The year index is expressed as a deviation from the mean $\bar{t}$.

[1] This group of four series was chosen because it exhibited large variation across series as well as marked trend and quarterly variation within series. This provided a good opportunity to test the performance of a variety of restricted least squares estimators.

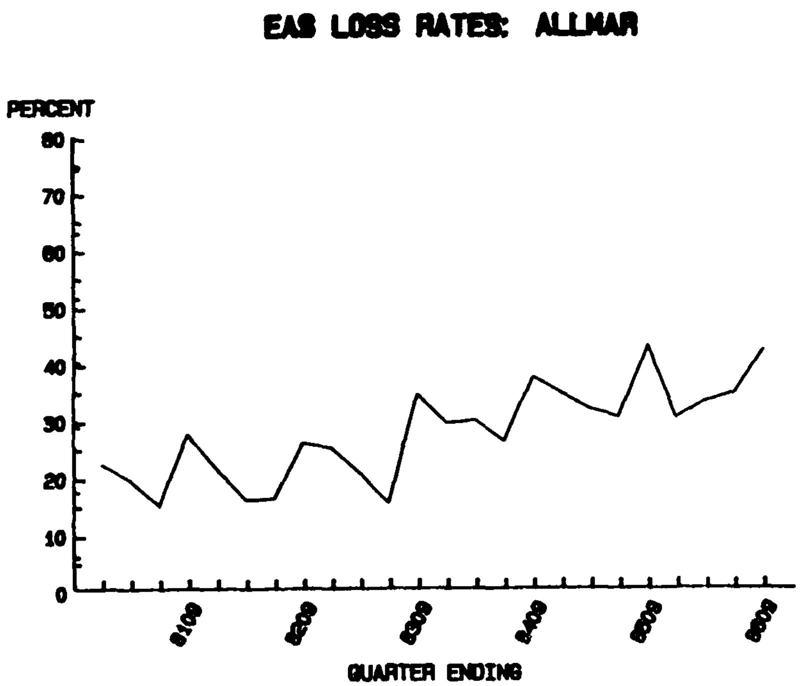

FIGURE 1a

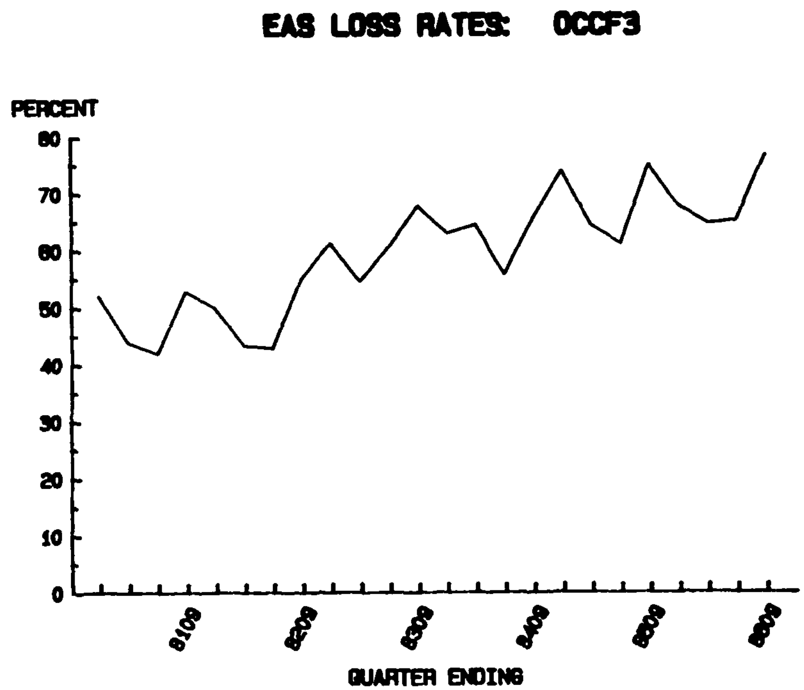

FIGURE 1b

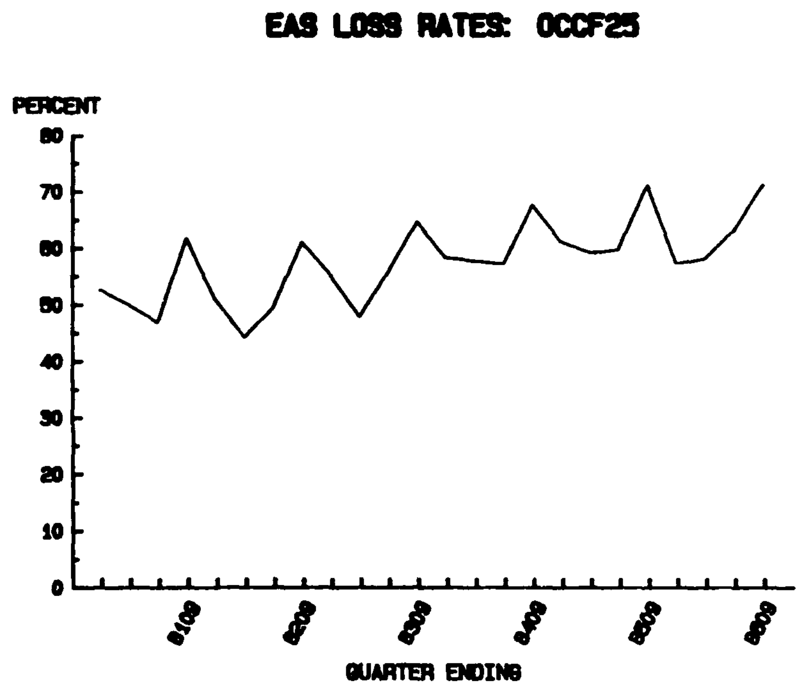

FIGURE 1c

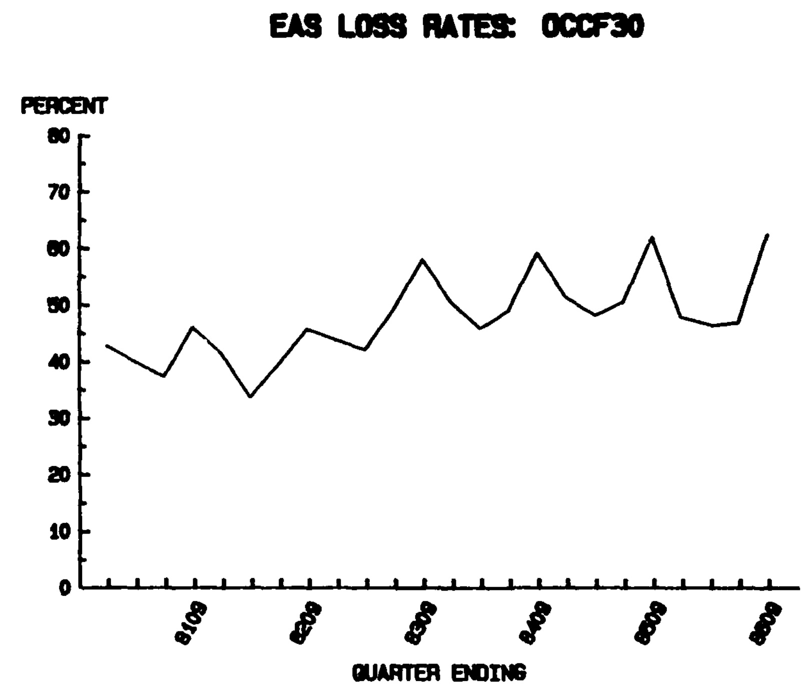

FIGURE 1d

The following multiple series specification was also considered:

$$Y_{itj} = \alpha_{ij} + \beta_i(t-\bar{t}) + \varepsilon_{itj} \tag{2}$$

The empirical Bayes estimation of parameters in (2) uses data from all series. This allows for interactions among several series. Here Y_{itj} is the actual rate for series i, year t, and quarter j. Again the ε_{itj}'s are assumed uncorrelated and homoscedastic. Specifications (1) and (2) differ only in their slope parameters. In the single series model, four distinct slopes β_j are allowed. In the multiple series model, the restriction of one slope β_i (for series i) common to all four quarters is imposed.

To generate forecasts using (1) or (2), the intercepts α and slopes β are first estimated from historical data. These estimates can then be inserted in the appropriate linear expression to yield one year ahead forecasts, quarter by quarter. For example, to forecast FY84 ALLMAR loss rates using (1), the 12 quarterly observations from FY81 through FY83 ($t = 1,2,3$ and $\bar{t} = 2$) are used to obtain estimates $\tilde{\alpha}_j$ and $\tilde{\beta}_j$ of the intercept and slope parameters. The forecasted loss rate for quarter j of FY84 then become:

$$\hat{Y}_{4j} = \tilde{\alpha}_j + \tilde{\beta}_j\ (4-2) = \tilde{\alpha}_j + 2\tilde{\beta}_j \tag{3}$$

where $t = 4$ corresponds to FY84. If (2) is used, the forecasted rate for series i and quarter j of FY84 is

$$\hat{Y}_{i4j} = \tilde{\alpha}_{ij} + 2\tilde{\beta}_i \qquad (4)$$

where the parameter estimates $\tilde{\alpha}_{ij}$ and $\tilde{\beta}_i$ are now allowed to depend on all 48 observations, i.e., 12 quarterly rates (FY81 through FY83) from each of the four series.

Parameter Estimation

A number of empirical Bayes estimators were considered in connection with both specifications (1) and (2). The basic form of an empirical Bayes estimator is derived as a weighted average of an unrestricted least squares estimator (ULSE) and a restricted least squares estimator (RLSE), where the weights are simple functions of an F-statistic testing the appropriateness of the restrictions. (See specifications A-1 and A-2 of Appendix A.)

For specification (1), the following sets of restrictions or hypotheses in the derivation of estimators were considered:

$$\begin{aligned} &H_A: \quad \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 \\ &H_B: \quad \beta_1 = \beta_2 = \beta_3 = \beta_4 \\ &H_{AB}: \quad \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4, \; \beta_1 = \beta_2 = \beta_3 = \beta_4 \end{aligned} \qquad (5)$$

The H_A and H_B restrictions imply no quarterly variation in intercepts and slopes, respectively. Imposing both sets of restrictions simultaneously, restrictions H_{AB}, implies one common line fits all four quarters.

The empirical Bayes estimator, E_{AB}, becomes

$$\tilde{\alpha}_j = \{1 - \frac{4}{6F}\} \hat{\alpha}_j + \{\frac{4}{6F}\} \hat{\alpha}$$

$$\tilde{\beta}_j = \{1 - \frac{4}{6F}\} \hat{\beta}_j + \{\frac{4}{6F}\} \hat{\beta} \; ; \qquad j = 1,2,3,4 \tag{6}$$

In (6), $\hat{\alpha}_j$ and $\hat{\beta}_j$ are the unrestricted least squares estimators (ULSE). The restricted least squares estimators under the restrictions H_A and H_B ($\hat{\alpha}$ and $\hat{\beta}$), are the simple averages of the $\hat{\alpha}_j$'s and $\hat{\beta}_j$'s. The value F is the F-statistic testing the restrictions H_{AB}.

The empirical Bayes estimator E_AE_B was also considered. It is specified as

$$\tilde{\alpha}_j = \{1 - \frac{1}{3F_1}\} \hat{\alpha}_j + \{\frac{1}{3F_1}\} \hat{\alpha}$$

$$\tilde{\beta}_j = \{1 - \frac{1}{3F_2}\} \hat{\beta}_j + \{\frac{1}{3F_2}\} \hat{\beta} \; ; \qquad j = 1,2,3,4 \tag{7}$$

Here (A-3) and (A-4) of Appendix A apply. The value F_1 is the F-statistic testing H_A and the value F_2 is the F-statistic testing H_B.

Two other specification (1) estimators were considered. The first is the restricted least squares estimator R_B associated with the restriction set H_B, where the intercepts are unrestricted ($\tilde{\alpha}_j = \hat{\alpha}_j$), but the slopes are constrained to be equal ($\tilde{\beta}_j = \hat{\beta}$). The second is the empirical Bayes estimator $E_A R_B$ with $\tilde{\beta}_j = \hat{\beta}$ and the intercepts determined by the first equation of (7).

Several similar multiple series (specification (2)) estimators were developed using the same notation as above. The sets of restrictions or hypotheses associated with the multiple series specification are

$$\begin{aligned} &H_A: \alpha_{1j} = \alpha_{2j} = \alpha_{3j} = \alpha_{4j} \quad (j = 1,2,3,4) \\ &H_B: \beta_1 = \beta_2 = \beta_3 = \beta_4 \\ &H_{AB}: \alpha_{1j} = \alpha_{2j} = \alpha_{3j} = \alpha_{4j} \quad (j = 1,2,3,4)\ , \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{and} \\ &\qquad \beta_1 = \beta_2 = \beta_3 = \beta_4 \end{aligned} \tag{8}$$

The H_A restrictions specify, for each quarter, equal intercepts across series. The H_B restrictions force a common slope for each series. The H_{AB} restrictions imply both H_A and H_B hold simultaneously, or four intercepts and one slope are common to all series.

The multiple series empirical Bayes estimator E_{AB} is defined as

$$\tilde{\alpha}_{ij} = \{1 - \frac{13}{15F}\} \hat{\alpha}_{ij} + \{\frac{13}{15F}\} \hat{\alpha}_j \tag{9}$$

$$\tilde{\beta}_i = \{1 - \frac{13}{15F}\} \hat{\beta}_i + \{\frac{13}{15F}\} \hat{\beta} \quad ; \quad \begin{array}{l} i = 1,2,3,4 \\ j = 1,2,3,4 \end{array}$$

In (9) $\hat{\alpha}_{ij}$ and $\hat{\beta}_i$ are the unrestricted least squares estimators (ULSE). The restricted estimators corresponding to H_A and H_B are, again, averages of the $\hat{\alpha}_{ij}$'s and $\hat{\beta}_i$'s ($\hat{\alpha}_j$ and $\hat{\beta}$). The F value, as in (6), is the F-statistic which tests the full set of restrictions H_{AB}.

The empirical Bayes estimator $E_A E_B$ for specification (2) is

$$\tilde{\alpha}_{ij} = \{1 - \frac{10}{12F_1}\} \hat{\alpha}_{ij} + \{\frac{10}{12F_1}\} \hat{\alpha}_j \tag{10}$$

$$\tilde{\beta}_i = \{1 - \frac{1}{3F_2}\} \hat{\beta}_i + \{\frac{1}{3F_2}\} \hat{\beta} \quad ; \quad \begin{array}{l} i = 1,2,3,4 \\ j = 1,2,3,\ 4 \end{array}$$

where F_1 and F_2 are the F-statistics testing H_A and H_B separately.

The final two estimators considered for specification (2) were, as in specification (1), denoted by R_B and $E_A R_B$. The estimator R_B is the restricted least squares estimator which requires the slopes to be equal ($\tilde{\beta}_i = \hat{\beta}$), but leaves the intercepts unrestricted ($\tilde{\alpha}_{ij} = \hat{\alpha}_{ij}$). The estimator $E_A R_B$ is the empirical Bayes estimator which shrinks the intercepts in accordance with the first equation of (10) and imposes the equal slopes restriction $\tilde{\beta}_i = \hat{\beta}$.

The kinds of differences that exist among the various parameter estimates are illustrated in Appendix B, Tables B-2 and B-3. These tables list all estimates associated with OccF3, for both specifications (1) and (2), estimated using FY83 through FY85 data. Notice that the ULSE estimates of "able B-3 are equivalent to the R_B estimates of Table B-2 since the multiple series specification (2) implies for each series a single slope across quarters.

The previous literature surrounding least squares and empirical Bayes methods compares these estimators through mean square error or, equivalently, square root of mean square error (SQRTMSE). This approach was also adopted here in assessing the forecasting performance of particular estimation techniques. Specifically, for each technique and series combination, the one year ahead forecasts (4 quarters) of FY84 through FY86 were generated using historical data, yielding 12 quarterly forecast errors. The SQRTMSE associated with these 12 errors was calculated as

$$SQRTMSE = \sqrt{\frac{e_1^2 + e_2^2 + \ldots + e_{12}^2}{12}} \qquad (11)$$

where the e's under the radical can denote the forecast errors corresponding to any technique applied to an individual series. Estimators that lead to small values of SQRTMSE are favored over those that lead to large values.

It should be mentioned that FY84 forecasts are based on parameter estimates generated by the FY81 through FY83 data. To achieve consistency, the forecasted loss

rates for FY85 and FY86 are also derived from parameter estimates that are functions of the most recent 3 years of data. Therefore, the FY85 forecasts depend on FY82 through FY84 values, while the FY86 forecasts are based on the FY83 through FY85 rates.

RESULTS

Table 1 lists the SQRTMSE's of all estimator-series combinations associated with the single series model. Note that for each series all three empirical Bayes estimators are more accurate than the ULSE. The smallest SQRTMSE in a series is given in bold. Furthermore, the best performer of the three, E_AR_B in each case, achieves percentage decreases in the SQRTMSE of the ULSE ranging from 18.38 percent for OccF30 to 40.51 percent for OccF3. Additionally, the estimator E_AE_B (double F) beats the estimator E_{AB} (single F) in three of the four series, the exception being OccF30 where their SQRTMSE's are approximately equal. Finally, although the restricted least squares estimator R_B does quite well, in three of the four series the empirical Bayes estimator E_AR_B ties or beats it, and in the other case (OccF30) the difference in SQRTMSE's is small.

Table 2 lists SQRTMSE's for the multiple series model. Here, the case for the empirical Bayes estimators is less favorable. In only three of the four series do all three empirical Bayes estimators achieve greater accuracy than the ULSE. Moreover, the maximum percentage reduction in the SQRTMSE of the ULSE attained by the empirical Bayes estimators ranges from only 4.58 percent for the ALLMAR series to 15.28 percent for the OccF3 series. Also, an empirical Bayes estimator is an overall winner for only two of the four series, as compared to three of four from Table 1. Finally, neither the double F empirical Bayes estimator E_AE_B nor the single F empirical Bayes estimator E_{AB} has a clear advantage over the other.

Table 1

SQRTMSE's
(Single Series Model)

Estimator	Series: ALLMAR	OccF3	OccF25	OccF30
Empirical Bayes:				
$E_A E_B$	.0391	.0493	.0478	.0455
E_{AB}	.0426	.0606	.0506	.0443
$E_A R_B$	**.0306**	**.0417**	**.0392**	.0404
ULSE	.0453	.0701	.0518	.0495
R_B	**.0306**	.0445	.0394	**.0400**

Table 2

SQRTMSE's
(Multiple Series Model)

Estimator	Series: ALLMAR	OccF3	OccF25	OccF30
Empirical Bayes:				
$E_A E_B$	.0298	**.0377**	.0391	.0486
E_{AB}	.0305	.0378	.0382	.0456
$E_A R_B$	.0292	.0396	**.0353**	.0519
ULSE	.0306	.0445	.0394	**.0400**
R_B	**.0286**	.0384	.0371	.0481

CONCLUSIONS

Confirming the theoretical evidence presented in the literature, empirical Bayes estimators outperformed least square estimators when applied to Marine Corps loss rate data. However, the double F estimator was not consistently more accurate than the single F estimator used in previous empirical Bayes analyses.

REFERENCES

Carter, G., & Rolph, J. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities. Journal of the American Statistical Association, 69, 880-885.

Casella, G. (1985). An introduction to empiric' Bayes data analysis. The American Statistician, 39, 83-87.

Efron B., & Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators. Part II: The empirical Bayes case. Journal of the American Statistical Association, 67, 130-139.

Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors: An empirical Bayes approach. Journal of the American Statistical Association, 68, 117-130.

Efron, B., & Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. Journal of the American Statistical Association, 70, 311-319.

Efron, B., & Morris, C. (1977). Stein's paradox in statistics. Scientific American, 236(5), 119-127.

Fay, R. E. III, & Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. Journal of the American Statistical Association, 74, 269-277.

Judge, G. G., & Bock, M. E. (1978). The statistical implications of pre-test and Stein-Rule estimators in econometrics, Amsterdam: North-Holland.

APPENDIX A

DERIVATION OF EMPIRICAL BAYES ESTIMATES IN THE LINEAR MODEL

DERIVATION OF EMPIRICAL BAYES ESTIMATES IN THE LINEAR MODEL

This appendix contains the basic setup for the general linear model and a brief development of empirical Bayes estimates in the linear model.

We assume the general linear model satisfies the following three criteria.

1. Normality: The vector Y (n x 1) has a multivariate Normal distribution and

$$Y \sim N(A_1\theta_1, C_1)$$

The matrix A_1 is n x k, has rank $k<n$ and the unknown parameter vector θ_1 is k x 1.

2. Orthogonality: The matrix A_1 is known and satisfies

$$A_1'A_1 = I_k$$

Where I_k is the k x k identity matrix.

3. I.i.d. errors: The matrix C_1 satisfies

$$C_1 = \sigma^2 I_n$$

where I_n is the n x n identity matrix and σ^2 is unknown.

Given this setup, the ordinary least squares estimate (OLS), $\hat{\theta}_1$, of the unknown parameter θ_1 is $\hat{\theta}_1 = A_1'Y$.

Empirical Bayes estimates are constructed by putting a prior distribution on the unknown parameter θ_1, then letting the observed data estimate a reasonable prior. In this paper, we use two different priors, one resulting in a single F test estimator and the second resulting in two F test estimators.
In both cases, the prior for θ_1 has the form:

$$\theta_1 \sim N(A_2\theta_2, C_2)$$

the prior parameter θ_2 is r x 1 and A_2 is k x r of rank r.

Single F Prior: This prior assumes that the prior errors are also iid, i.e.

$$C_2 = \tau^2 I_k$$

Two F Prior: This prior assumes that:

$$\theta_1 = \begin{pmatrix} \theta_{11} \\ \theta_{12} \end{pmatrix}$$

i.e. θ_1 can be broken into components, such that θ_{11} is k_1 x 1 and θ_2 is k_2 x 1, $k_1+k_2=k$.

$$A_2 = \begin{pmatrix} A_{21} & 0 \\ 0 & A_{22} \end{pmatrix}$$

where A_{21}, k_1 x r_1, and A_{22}, k_2 x r_2, are known, $r_1 + r_2 = r$.

θ_2 can also be broken into components,

$$\theta_2 = \begin{pmatrix} \theta_{21} \\ \theta_{22} \end{pmatrix}$$

with θ_{21} r_1 x 1 and θ_{22} r_2 x 1.

Finally, the covariance matrix

$$C_2 = \begin{pmatrix} \tau_1^2 I_{k_1} & 0 \\ 0 & \tau_2^2 I_{k_1} \end{pmatrix}$$

The two F tests can trivially be extended to arbitrary numbers of F tests if the prior is suitably split into components.

Bayes Estimates: Under quadratic loss, the Bayes estimate of θ_1 is just the posterior mean of θ_1 given the observed data Y.

In the one F case, this is just

$$(1-W)\hat{\theta}_1 + WA_2\theta_2$$

where $W = \sigma^2 / (\tau^2+\sigma^2)$ and $\hat{\theta}_1$ is the OLS estimate.

Thus, the Bayes estimate is just a weighted average of the OLS estimate θ_1 and the prior mean $A_2\theta_2$. Note that the assumption $E(\theta_1) = a_2\theta_2$ can also be written as a restriction on θ_1, i.e.

$$R\ \theta_1 = 0$$

In the two F case, the Bayes estimate is similar, with separate weights for each component. Thus, the Bayes estimate is

$$(1-W_i)\hat{\theta}_{1i} + W_i A_{2i}\theta_{2i} \quad i=1,2$$

where

$$W_i = \sigma^2 / (\tau_i^2 + \sigma^2)$$

and $\hat{\theta}_{1i}$ i=1,2 are the OLS estimates of θ_{1i} i=1,2.

Empirical Bayes Estimates: To derive empirical Bayes estimates, we assume the prior parameters are unknown, except for a_2, and try to estimate them. Under the Bayesian setup, Y has a Normal distribution with mean $A_1 A_2 \theta_2$. Therefore,

$$\hat{\theta}_2 = (A_2' A_2)^{-1} A_2' A_1' Y$$

is the OLS unbiased estimate of θ_2.

In the two F case, $\hat{\theta}_2$ will separate into component OLS estimators $\hat{\theta}_{21}$ and $\hat{\theta}_{22}$.

Using $\hat{\theta}_1$ and $\hat{\theta}_2$, we can calculate several quadratic forms, which will be used to estimate σ^2, and τ^2 (or τ_1^2 and τ_2^2 in the two F case).

First, define

$$RSS_0 = (Y - A_1\hat{\theta}_1)'(Y - A_1\hat{\theta}_1)$$

This is just the familiar error sum of squares.

Next we define

$$RSS_{tot} = (Y - A_1 A_2 \hat{\theta}_2)'(Y - A_1 A_2 \hat{\theta}_2)$$

and

$$\begin{aligned} RSS &= RSS_{tot} - RSS_0 \\ &= (\hat{\theta}_1 - A_2\hat{\theta}_2)'(\hat{\theta}_1 - A_2\hat{\theta}_2) \end{aligned}$$

Now, RSS and RSS_0 are the numerator and denominator sum of squares in the usual F test of the hypothesis, $\theta_1 = A_2\theta_2$.

Using standard linear model methods, it is easy to show

$$RSS_0 \sim \sigma^2 \chi^2(n-k)$$

where $\chi^2(n)$ is a central χ^2 distribution with n degrees of freedom.

The distribution of RSS, however, depends on which prior is used.

For the one F prior, i.e. the one with the iid errors with variance τ^2,

$$RSS \sim (\sigma^2+\tau^2)\chi^2(k-r)$$

However, for the two F prior, RSS splits into two components, $RSS_1 + RSS_2$, with

$$RSS_1 = = (\hat{\theta}_{11}-A_{21}\hat{\theta}_{21})'(\hat{\theta}_{11}-A_{21}\hat{\theta}_{21})$$

and

$$RSS_2 = = (\hat{\theta}_{12}-A_{22}\hat{\theta}_{22})'(\hat{\theta}_{12}-A_{22}\hat{\theta}_{22})$$

and

$$RSS_i \sim (\sigma^2+\tau_i^2)\chi^2(k_i-r_i) \quad i = 1,2$$

In the one F case, if we let

$$\hat{W} = \frac{RSS_0/n-k}{RSS/k-r-2} \tag{A-1}$$

then $\hat{W}$ is an unbiased estimate of W. and, hence,

$$(1-\hat{W})\hat{\theta}_1+\hat{W}A_2\hat{\theta}_2 \tag{A-2}$$

is an empirical Bayes estimate of θ_1. But

$$\hat{W} = \frac{k-r-2}{k-r}F^{-1}$$

so that the estimator

$$(1-F^{-1})\hat{\theta}_1+F^{-1}A_2\hat{\theta}_2$$

is also an empirical Bayes estimate of θ_1, although F^{-1} is not an unbiased estimate of W.

In the two F case, the situation is analogous.

$$\hat{W}_1 = \frac{RSS_0/n-k}{RSS_1/k_1-r_1-2}$$

and

$$\hat{W}_2 = \frac{RSS_0/n-k}{RSS_2/k_2-r_2-2}$$

are unbiased estimates of W_1 and W_2 resp.. Algebraically, the weights are equivalent to:

$$\hat{W}_i = \frac{k_i-r_i-2}{k-r}F_i^{-1} \quad i=1,2 \tag{A-3}$$

and the 2-F empirical Bayes estimate is:

$$(1-\hat{W})\hat{\theta}_1+\hat{W}A_2\hat{\theta}_2 \tag{A-4}$$

In the 2-F case, $F_i\, i=1,2$ are the usual components of variance F test of the hypotheses

$$H_i : \theta_{1i} = A_{2i}\theta_{2i} \quad i=1,2$$

In both the one F and two F priors, the F test provides a plausible way to shrink a least squares estimate toward a restricted estimator based on a submodel of the original linear model.

APPENDIX B

SUPPORTING TABLES

Table B-1

Quarterly EAS Loss Rates - FY81 Through FY86

Quarter Ending	ALLMAR	OccF3	OccF25	OccF30
8012	.4918	.5201	.5261	.4270
8103	.4668	.4368	.4994	.3987
8106	.4280	.4163	.4675	.3715
8109	.5383	.5280	.6165	.4610
8112	.4825	.4975	.5070	.4136
8203	.4362	.4315	.4424	.3356
8206	.4392	.4271	.4951	.3971
8209	.5243	.5497	.6104	.4584
8212	.5163	.6117	.5505	.4392
8303	.4761	.5441	.4777	.4209
8306	.5312	.6054	.5565	.4943
8309	.5990	.6778	.6457	.5806
8312	.5546	.6285	.5816	.5037
8403	.5594	.6438	.5761	.4575
8406	.5260	.5546	.5713	.4906
8409	.6258	.6564	.6755	.5920
8412	.6003	.7384	.6098	.5124
8503	.5759	.6422	.5912	.4814
8506	.5632	.6093	.5974	.5054
8509	.6737	.7490	.7120	.6197
8512	.5620	.6759	.5717	.4784
8603	.5877	.6446	.5805	.4642
8606	.6014	.6498	.6307	.4694
8609	.6668	.7653	.7116	.6250

Table B-2

Estimates Generating OccF3 FY86 Forecasts
(Single Series Models)

Parameter	ULSE	$E_A E_B$	E_{AB}	R_B	$E_A R_B$
		Estimator			
$\tilde{\alpha}_1$	.6595	.6577	.6533	.6595	.6577
$\tilde{\alpha}_2$	.6100	.6125	.6184	.6100	.6125
$\tilde{\alpha}_3$	.5898	.5941	.6041	.5898	.5941
$\tilde{\alpha}_4$	.6944	.6894	.6779	.6944	.6894
$\tilde{\beta}_1$	.0634	.0521	.0557	.0375	.0375
$\tilde{\beta}_2$	.0491	.0440	.0456	.0375	.0375
$\tilde{\beta}_3$	.0020	.0174	.0124	.0375	.0375
$\tilde{\beta}_4$	.0356	.0364	.0362	.0375	.0375

Table B-3

Estimates Generating OccF3 FY86 Forecasts
(Multiple Series Model)

Parameter	Estimator ULSE	$E_A E_B$	E_{AB}	R_B	$E_A R_B$
$\tilde{\alpha}_1$	.6595	.6546	.6532	.6595	.6546
$\tilde{\alpha}_2$	.6100	.6060	.6048	.6100	.6060
$\tilde{\alpha}_3$	.5898	.5876	.5870	.5898	.5876
$\tilde{\alpha}_4$	.6944	.6920	.6913	.6944	.6920
$\tilde{\beta}$	.0375	.0348	.0372	.0329	.0329

DISTRIBUTION LIST

Commandant of the Marine Corps (MA, MPP-20, MPI-40)
Commander, Air Force Human Resources Laboratory (Manpower and Personnel Division), Brooks Air Force Base, TX
Commander, U.S. Army Research Institute for the Behavioral and Social Sciences (PERI-POT-I), Alexandria, VA
Chief of Naval Operations (OP-135)
Superintendent, Naval Postgraduate School
Center for Naval Analyses
Defense Technical Information Center (DTIC) (2)